Table 3 shows MAE metric of all tested measures about all *r* = 0.7, 0.9 within estimation process. The last column shows average MAE metrics over all values of *r* and shaded cells indicate best values. As a convention, we define that preeminent measures (dominant measures) are ones in top-5 lists.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *r*=0.7  *k*=5 | *r*=0.7  *k*=20 | *r*=0.7  *k*=50 | *r*=0.7  *k*=100 | *r*=0.9  *k*=5 | *r*=0.9  *k*=20 | *r*=0.9  *k*=50 | *r*=0.9  *k*=100 | Average  (MAE) |
| Cosine | 0.6364 | 0.6368 | 0.6370 | 0.6367 | 0.6837 | 0.6845 | 0.6868 | 0.6846 | 0.6608 |
| CON | 0.6457 | 0.6460 | 0.6466 | 0.6460 | 0.6923 | 0.6939 | 0.6970 | 0.6945 | 0.6703 |
| Pearson | 0.6591 | 0.6602 | 0.6611 | 0.6599 | 0.7073 | 0.7087 | 0.7079 | 0.7064 | 0.6838 |
| MSDJ | 0.6363 | 0.6366 | 0.6369 | 0.6365 | 0.6835 | 0.6847 | 0.6868 | 0.6847 | 0.6608 |
| NHSM | 0.6375 | 0.6378 | 0.6380 | 0.6378 | 0.6872 | 0.6889 | 0.6909 | 0.6886 | 0.6633 |
| PIP | 0.6365 | 0.6373 | 0.6374 | 0.6371 | 0.6951 | 0.6944 | 0.6977 | 0.6953 | 0.6664 |
| TA | 0.6362 | 0.6365 | 0.6368 | 0.6364 | 0.6833 | 0.6843 | 0.6865 | 0.6841 | 0.6605 |
| TAJ | 0.6364 | 0.6366 | 0.6369 | 0.6366 | 0.6838 | 0.6853 | 0.6874 | 0.6851 | 0.6610 |
| SMD | 0.6419 | 0.6405 | 0.6415 | 0.6412 | 0.6803 | 0.6846 | 0.6869 | 0.6849 | 0.6627 |

**Table 3.** MAE metric within estimation process

Top-3 measures according to MAE metric within estimation process are TA, MSDJ, and cosine whose average MAE metrics are 0.6605, 0.6608, and 0.6608, respectively. Our TA measure is in the top-3 list given MAE metric. Shortly, dominant orders of our measures TA, TAJ, and SMD are 1st, 4th, and 5th among 9 measures given MAE metric, respectively.

Table 4 shows precision metric of all tested measures about all *r* = 0.7, 0.9 within recommendation process given precision metric. The last column shows average precision metrics over all values of *r* and shaded cells indicate best values.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *r*=0.7  *k*=5 | *r*=0.7  *k*=20 | *r*=0.7  *k*=50 | *r*=0.7  *k*=100 | *r*=0.9  *k*=5 | *r*=0.9  *k*=20 | *r*=0.9  *k*=50 | *r*=0.9  *k*=100 | Average  (Precision) |
| Cosine | 0.0258 | 0.0256 | 0.0253 | 0.0255 | 0.1849 | 0.1867 | 0.1853 | 0.1897 | 0.1061 |
| CON | 0.0262 | 0.0261 | 0.0258 | 0.0260 | 0.1875 | 0.1895 | 0.1884 | 0.1925 | 0.1078 |
| Pearson | 0.0275 | 0.0273 | 0.0270 | 0.0272 | 0.2077 | 0.2096 | 0.2088 | 0.2131 | 0.1185 |
| MSDJ | 0.0258 | 0.0257 | 0.0254 | 0.0256 | 0.1850 | 0.1867 | 0.1853 | 0.1898 | 0.1062 |
| NHSM | 0.0258 | 0.0257 | 0.0254 | 0.0256 | 0.1844 | 0.1861 | 0.1846 | 0.1891 | 0.1058 |
| PIP | 0.0260 | 0.0258 | 0.0255 | 0.0257 | 0.1847 | 0.1866 | 0.1850 | 0.1895 | 0.1061 |
| TA | 0.0258 | 0.0256 | 0.0253 | 0.0255 | 0.1850 | 0.1867 | 0.1853 | 0.1897 | 0.1061 |
| TAJ | 0.0258 | 0.0257 | 0.0254 | 0.0256 | 0.1850 | 0.1866 | 0.1852 | 0.1896 | 0.1061 |
| SMD | 0.0143 | 0.0142 | 0.0142 | 0.0142 | 0.0396 | 0.0400 | 0.0396 | 0.0398 | 0.0270 |

**Table 4.** Precision metric within recommendation process

Top-3 measures according to precision metric within recommendation process are Pearson, CON, and MSDJ whose average precision metrics are 0.1185, 0.1078, and 0.1062, respectively. Shortly, dominant orders of our measures TA, TAJ, and SMD are 4th, 5th, and 9th among 9 measures given precision metric, respectively.

Table 5 shows recall metric of all tested measures about all *r* = 0.7, 0.9 within recommendation process given recall metric. The last column shows average recall metrics over all values of *r* and shaded cells indicate best values.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *r*=0.7  *k*=5 | *r*=0.7  *k*=20 | *r*=0.7  *k*=50 | *r*=0.7  *k*=100 | *r*=0.9  *k*=5 | *r*=0.9  *k*=20 | *r*=0.9  *k*=50 | *r*=0.9  *k*=100 | Average  (Recall) |
| Cosine | 0.8512 | 0.8531 | 0.8525 | 0.8523 | 0.7406 | 0.7378 | 0.7379 | 0.7383 | 0.7955 |
| CON | 0.8502 | 0.8525 | 0.8515 | 0.8516 | 0.7353 | 0.7328 | 0.7329 | 0.7330 | 0.7925 |
| Pearson | 0.8494 | 0.8510 | 0.8504 | 0.8505 | 0.7192 | 0.7171 | 0.7196 | 0.7186 | 0.7845 |
| MSDJ | 0.8514 | 0.8529 | 0.8527 | 0.8524 | 0.7406 | 0.7378 | 0.7380 | 0.7385 | 0.7955 |
| NHSM | 0.8492 | 0.8506 | 0.8506 | 0.8503 | 0.7360 | 0.7333 | 0.7333 | 0.7339 | 0.7922 |
| PIP | 0.8539 | 0.8549 | 0.8544 | 0.8543 | 0.7373 | 0.7359 | 0.7355 | 0.7362 | 0.7953 |
| TA | 0.8513 | 0.8530 | 0.8526 | 0.8523 | 0.7406 | 0.7374 | 0.7376 | 0.7382 | 0.7954 |
| TAJ | 0.8511 | 0.8525 | 0.8524 | 0.8521 | 0.7398 | 0.7367 | 0.7370 | 0.7374 | 0.7949 |
| SMD | 0.8757 | 0.8784 | 0.8772 | 0.8772 | 0.8065 | 0.7996 | 0.8008 | 0.8007 | 0.8395 |

**Table 5.** Recall metric within recommendation process

Top-3 measures according to recall metric within recommendation process are SMD, MSDJ, and cosine whose average recall metrics are 0.8395, 0.7955, and 0.7955, respectively. Our SMD measure is in the top-3 list given recall metric. Shortly, dominant orders of our measures TA, TAJ, and SMD are 4th, 6nd, and 1st among 9 measures given recall metric, respectively.

From metrics MAE, precision, and recall shown in tables 3, 4, 5, respectively, it is not easy to determine which measures are the best. Remind that estimation process is evaluated by MAE metric and recommendation process is evaluated by both precision metric and recall metric. F1 metric is the way to assembling precision and recall together. F1 metric is specified by following equation (Herlocker, Konstan, Terveen, & Riedl, 2004, p. 25). The larger F1 is, the better measures are.

Shortly, MAE is used to evaluate estimation process and F1 is used to evaluate recommendation process. Table 6 which is derived from tables 3, 4, and 5 shows average MAE values and F1 values of all measures. Shaded cells indicate best values.

|  |  |  |
| --- | --- | --- |
|  | MAE | F1 |
| Cosine | 0.6608 | 0.207723 |
| CON | 0.6703 | 0.189794 |
| Pearson | 0.6838 | 0.187325 |
| MSDJ | 0.6608 | 0.187244 |
| NHSM | 0.6633 | 0.187242 |
| PIP | 0.6664 | 0.187211 |
| TA | 0.6605 | 0.187144 |
| TAJ | 0.6610 | 0.186727 |
| SMD | 0.6627 | 0.052180 |

**Table 6.** General MAE and F1 over all measures

Top-3 measures according to F1 metric within recommendation process are cosine, CON, and Pearson whose average recall metrics are 0.207723, 0.189794, and 0.187325. Shortly, dominant orders of our measures TA, TAJ, and SMD are 7th, 8th, and 9th among 9 measures given F1 metric, respectively.

Figure 2 which is extracted from table 6 draws comparison over all measures regarding MAE metric.

Chart, bar chart

Description automatically generated

**Figure 2.** Measure comparison with MAE metric

Figure 3 which is extracted from table 6 draws comparison over all measures regarding F1 metric.

A picture containing table

Description automatically generated

**Figure 3.** Measure comparison with F1 metric

It is easy to recognize that table 6 is the general evaluation of all measures regarding estimation process and recommendation process. We cannot unify MAE and F1 as we unify precision and recall because estimation process and recommendation process are not always proportional. Therefore, let *A* be set of top-3 measures regarding MAE and let *B* be set of top-5 measures regarding F1. The intersection of *A* and *B* contains best measures. From table 6, we have *A* = {TA, MSDJ, cosine} and *B* = {cosine, CON, Pearson}. Obviously, the best measure in general comparison is cosine.

Although it is totally possible to evaluate measures with two metrics MAE and F1, it is better to go further with other metrics. For estimation process, some popular metrics which are different from MAE are mean squared error (MSE) and correlation coefficient (R). MSE is calculated by equation 50 (Herlocker, Konstan, Terveen, & Riedl, 2004, p. 21) in which *n* is the total number of estimated items while *vj*’ and *vj* are predictive rating and true rating of item *j*, respectively. Of course, we have predictive vector *v*’ and true vector (tested vector) *v*.

The smaller MSE is, the more accurate the measures are and so the better the algorithm is. R is used to evaluate correlation between predictive vector *v*’ and true vector *v*. It is really Pearson correlation. The larger R is, the better the measures are. R metric is specified by following equation (Montgomery & Runger, 2010, p. 432).

Where and are mean values of tested item and predictive item, respectively.

Table 7 shows MSE metric of all tested measures about all values of *r*. Shaded cells indicate best values.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *r*=0.7  *k*=5 | *r*=0.7  *k*=20 | *r*=0.7  *k*=50 | *r*=0.7  *k*=100 | *r*=0.9  *k*=5 | *r*=0.9  *k*=20 | *r*=0.9  *k*=50 | *r*=0.9  *k*=100 | Average  (MSE) |
| Cosine | 0.7073 | 0.7095 | 0.7082 | 0.7074 | 0.8179 | 0.8187 | 0.8229 | 0.8190 | 0.7639 |
| CON | 0.7183 | 0.7194 | 0.7190 | 0.7173 | 0.8207 | 0.8241 | 0.8297 | 0.8253 | 0.7717 |
| Pearson | 0.7288 | 0.7310 | 0.7320 | 0.7290 | 0.8240 | 0.8259 | 0.8250 | 0.8227 | 0.7773 |
| MSDJ | 0.7068 | 0.7093 | 0.7078 | 0.7070 | 0.8148 | 0.8160 | 0.8200 | 0.8161 | 0.7622 |
| NHSM | 0.7091 | 0.7119 | 0.7103 | 0.7095 | 0.8231 | 0.8251 | 0.8289 | 0.8244 | 0.7678 |
| PIP | 0.7044 | 0.7076 | 0.7062 | 0.7051 | 0.8327 | 0.8301 | 0.8372 | 0.8319 | 0.7694 |
| TA | 0.7054 | 0.7074 | 0.7062 | 0.7053 | 0.8131 | 0.8142 | 0.8180 | 0.8136 | 0.7604 |
| TAJ | 0.7061 | 0.7084 | 0.7069 | 0.7061 | 0.8138 | 0.8155 | 0.8193 | 0.8151 | 0.7614 |
| SMD | 0.7203 | 0.7183 | 0.7190 | 0.7184 | 0.8215 | 0.8283 | 0.8332 | 0.8299 | 0.7736 |

**Table 7.** MSE metric within estimation process

Top-3 measures according to MSE metric within estimation process are TA, TAJ, and MSDJ whose average MSE metrics are 0.7604, 0.7614, and 0.7622. Our measures TA and TAJ is in the top-3 list given MSE metric. Shortly, dominant orders of our measures TA, TAJ, and SMD are 1st, 2nd, and 8th among 9 measures given MSE metric, respectively.

Table 8 shows R metric of all tested measures about all values of *r*. Shaded cells indicate best values.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *r*=0.7  *k*=5 | *r*=0.7  *k*=20 | *r*=0.7  *k*=50 | *r*=0.7  *k*=100 | *r*=0.9  *k*=5 | *r*=0.9  *k*=20 | *r*=0.9  *k*=50 | *r*=0.9  *k*=100 | Average  (R) |
| Cosine | 0.1938 | 0.1959 | 0.1963 | 0.1976 | 0.0962 | 0.1036 | 0.1030 | 0.1032 | 0.1487 |
| CON | 0.1671 | 0.1584 | 0.1545 | 0.1579 | 0.0545 | 0.0551 | 0.0535 | 0.0537 | 0.1068 |
| Pearson | 0.0904 | 0.0828 | 0.0807 | 0.0834 | 0.0253 | 0.0212 | 0.0187 | 0.0201 | 0.0528 |
| MSDJ | 0.1998 | 0.1988 | 0.1995 | 0.2012 | 0.0948 | 0.1042 | 0.1024 | 0.1031 | 0.1505 |
| NHSM | 0.1992 | 0.1973 | 0.1981 | 0.2002 | 0.0923 | 0.0974 | 0.0979 | 0.0988 | 0.1477 |
| PIP | 0.2005 | 0.1978 | 0.1980 | 0.1999 | 0.0840 | 0.0919 | 0.0901 | 0.0916 | 0.1442 |
| TA | 0.1986 | 0.2001 | 0.2004 | 0.2022 | 0.0979 | 0.1053 | 0.1043 | 0.1056 | 0.1518 |
| TAJ | 0.2014 | 0.2001 | 0.2013 | 0.2033 | 0.0971 | 0.1047 | 0.1025 | 0.1036 | 0.1518 |
| SMD | 0.2048 | 0.2087 | 0.2059 | 0.2064 | 0.1621 | 0.1628 | 0.1656 | 0.1662 | 0.1853 |

**Table 8.** R metric within estimation process

Top-3 measures according to R metric within estimation process are SMD, TA, and TAJ whose average R metrics are 0.1853, 0.1518, and 0.1518. Our measures SMD, TA, and TAJ are in the top-3 list given R metric. Shortly, dominant orders of our measures SMD, TA, and TAJ are 1st, 2nd, and 3rd among 9 measures given R metric, respectively.

The best measure TAJ drawn from tables 6 are still in lists of top-5 measures with regard to MSE and R shown in tables 7 and 8. This implies the same semantics of MAE, MSE, and R within estimation process. It is possible to conclude that the important problem is to split the evaluation process into two sub-processes such as estimation and recommendation. For each sub-process, we only need to choose one representative metric. In this research, we choose MAE and F1 as representative metrics for estimation process and recommendation process, respectively.

Although the best measure is cosine with representative metrics MAE and F1, three our measures TA, TAJ and SMD are also good measures. As usual, we define that preeminent measures (dominant measures) are ones in top-3 lists. It is useful to compare cosine, TA, TAJ, and SMD but it is impossible to unify metrics MAE, MSE, and R together. However, we can compare them by taking advantages of some transformations. Let I-R be inverse of R metric. Let I-Precision be inverse of precision metric and let I-Recall be inverse of recall metric. The smaller I-R, I-Precision, and I-Recall are, the better the measures are. I-R, I-Precision, and I-Recall are specified by following equation. Hence, I-R, I-Precision, and I-Recall are replacers of R, Precision, and Recall.

Table 9 lists metrics MAE, MSE, I-R, I-Precision, and I-Recall of preeminent measures NHSM, SMD, and TAJ.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | MAE | MSE | I-R | I-Precision | I-Recall |
| Cosine | 0.6608 | 0.7639 | 0.8513 | 0.8939 | 0.2045 |
| TA | 0.6605 | 0.7604 | 0.8482 | 0.8939 | 0.2046 |
| TAJ | 0.6610 | 0.7614 | 0.8483 | 0.8939 | 0.2051 |
| SMD | 0.6627 | 0.7736 | 0.8147 | 0.9730 | 0.1605 |

**Table 9.** Comparison of cosine, TA, TAJ, and SMD with MAE, MSE, I-R, I-Precision, and I-Recall

From table 9, TA is the best with MAE and MSE. SMD is the best with I-R, and I-Recall. Cosine, TA, and TAJ are the best with I-Precision.

Figure 4 shows radar 3D chart of preeminent measures cosine, TA, TAJ, and SMD regarding MAE, MSE, I-R, I-Precision, and I-Recall.

Chart, bar chart

Description automatically generated

**Figure 4.** Comparison of cosine, TA, TAJ, and SMD with MAE, MSE, I-R, I-Precision, and I-Recall

**References**

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53. doi:10.1145/963770.963772

Montgomery, D. C., & Runger, G. C. (2010). *Applied Statistics and Probability for Engineers* (5th ed.). Hoboken, New Jersey, USA: John Wiley & Sons. Retrieved September 6, 2016, from https://books.google.com.vn/books?id=\_f4KrEcNAfEC